Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 4347 – 4351
Procedia Engineering
www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

Optimal partial postponement capacity of the single-period product under stochastic demand

Yanhong QIN [a]*

School of Management, Chongqing Jiaotong University, Chongqing, 400074, China

Abstract

Many companies have adopted the postponement strategy to manage their supply chain and achieve the continuous competitive advantage. There have been many literatures solving the postponement degree of product manufacture, such as Graman (2010), and our paper is to solve the similar problem. The target of our paper is to a new cost model to solve the partial postponement problem by adding penalty cost parameter of shortage under stochastic demand based on the research work by Graman (2010) in the European Journal of Operational Research.

Keywords: Postponement; Inventory; Mass customization

1. Introduction

Postponement strategy is an effective strategy to achieve mass customization and it has been perceived as one of the major supply chain management practices, because it can better deal with the product proliferation without incurring large operating costs caused by postponement activities (Li et al., 2006). There have been many researches on the postponement problem, typically, as in Graman and Magazine (2002, denoted G&M(2002)), the authors studied the impact of postponement capacity on the benefits of inventory savings, given a defined customer service level, denoted as fill rate, which is decided by

* Corresponding author. Tel.:+0086-23-62652486
E-mail address:qinyanhong24@163.com.

doi:10.1016/j.proeng.2011.08.816

managers in the paper, and then a model of single-period and two products capacitated –postponement inventory is analyzed, where the non-postponed finished final product inventory and the generic product inventory (i.e. postponed inventory, which will be customized until the real customer demand is known) are both held by the manufacture. The two final different products can be made from the generic products (i.e. postponed products) by packaging, additional parts or other customized service etc. to meet different demand. The finished product inventory will be used to meet the real demand first, but once the demand are excess of finished product inventory, some or all of the generic product will be completed to meet the excess demand as much as possible within the specified delivery time. The author obtained some important conclusions, such as when the fill rate, the coefficient of variation, the number of products being postponed increase, the demands are more negatively correlated and the demand distribution of different products are approaching each other, and then inventory saving will increase and most importantly, the author observed an important phenomenon that a relatively small amount of postponement capacity (about 40% of total expected demand) can achieve all of the benefits of completely postponing all demands, i.e. keeping all the inventory as generic product and different customized parts rather than any final finished products, and the customized parts can be assembled to the generic product to form the final customized products quickly until the real demands is certain. This important phenomenon will inspire many firms to adopt the partial postponement strategy, and especially, postponement strategy is relative to some additional postponement cost, such as the investment cost, processing cost, handling cost of common inventory. But there is a potential preliminary behind the numerical analysis and observation, i.e. the fill rates of different products are the same.

So in the later research of Graman(2010, denoted G(2010)), the author further set a similar model to the model in G&M(2002), but there are some distinct difference. The first difference is G(2010) focused on solving the minimum-cost objective function by considering some additional costs caused by postponement, where the postponed manufacturing or assembly will cause the more frequent setups of production line to process smaller lot size, additional handling, packaging to facilitate the handling and maintaining integration of generic products, denoted as assemble labor and material cost. Secondly, G(2010) showed some additional different conclusions based on the research of G&M(2002), such that when the value of generic products and relative postponement cost decrease (including packaging postponed inventory to maintain product integrity, additional operation, waiting caused by inability to delivery within the specified lead time), the holding cost increase, etc., the total inventory and expected total cost will decrease. The third difference between the two continuous researches is: in G&M(2002), the assumption that each fill rate or the expected number of stockouts for each customized final products equal is the basis for comparison among different level of partial or capacitated postponement strategy and other sensitivity analysis of different parameters change on the total inventory level of generic products and final products. But in the version of G(2010) the fill rate or the expected number of stockouts is set as an constraint in solving the minium cost non-linear programing problem, and in the process of reasoning and analysis, G(2010) didn't discriminate the regions [8][9] and [7] in Fig 1. That is to say, G (2010) treat the condition happened in region [8] and [9] as the same in region [7] in the computating the expected number of stockouts, as shown in the equation (C.4) and (C.5) in the appendix C in G (2010), which will cause error reasoning in solving the non-programming problem. Importantly, the fill rates for different customer are often different according to the profits obtained from the sales for different customer, the important degree of the orders, or the penalty cost caused by the demand unmet. The more important of the order, the more profit obtained from the order or the bigger penalty cost, the responding order should be met in the more anterior sequence, which is a universal phenomenon in many enterprises, so there should be some difference between the fill rates for different customer or products, and we will demonstrate that the problem formulation and computation will be simplified and direct by introducing the penalty cost parameter substituting for the fill rate.

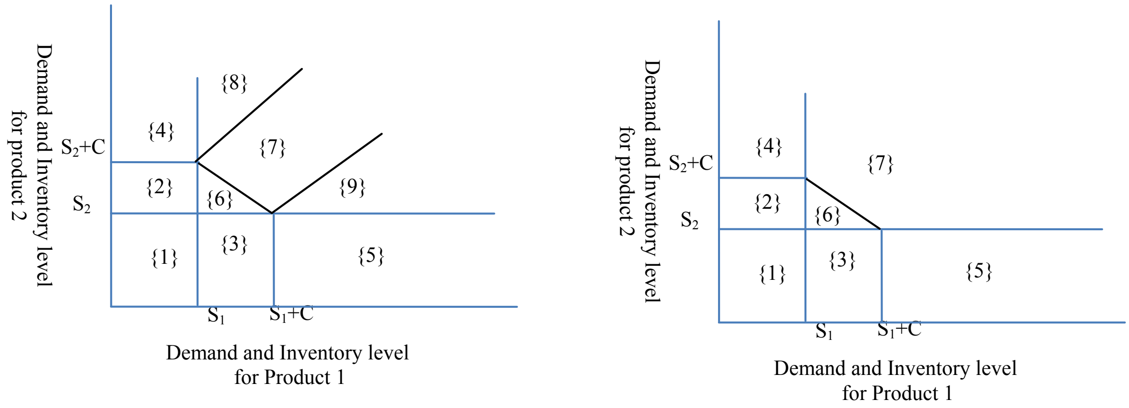

Fig.1. (a) Graphical depiction of inventory levels and regions for a partial postponement scenario where realizations of demand and stockouts can occur, as same in G&M(2002); (b) Graphical depiction of inventory levels defining regions where realizations of demand can occur for a partial-postponement scenario, as same in G(2010).

So the main targets of our paper include two aspects: (1) why the expression of expected stockout product should consider the region [7], [8] and [9] differently, and discriminate the integration expression of expected number of stockouts happened in region [7], [8] and [9]; (2) when the introduce of the different penalty cost parameter rather than the equal fill rate parameter for different customized product, and when the shortage cost is set as an item in the objective total cost function but not as a constaint included in G(2010), the reasoning and computation can be simplified, and more importantly, it is more fit the decision condition as many companies meet. For convenience and comparation, the following notation and definitions are used as those in G(2010) except the parameter of penalty cost t_i and inventory number of left over LO_i .

2. Expression for the expected stockouts of inventory

In model of G(2010), the objective function was to minimize the cost made up of assembly cost, postponement cost, packaging cost and holding cost of combined finished goods with postponed inventory. The constraints include fill rate constraints, boundary condition constraints, postponement capacity allocation constraints and non-negativity constaints. The expression of the expected stockouts $E[SO_i]$ ($i=1,2$) in the region [1]-[5] is same to G&M(2002) and G(2010), when the real demand of both products can't be met by all finished product inventory and generic or postponed inventory which is happened in region [7]-[9] of fig.1(a) or in region [7] of fig.1(b), then the computation of expected stockouts for each product is based on the decision rule: equalize the fill rates of each product (P38, in G(2010)), i.e. $\frac{S_1+P_1}{x_1}=\frac{S_2+P_2}{x_2}$, and $P_1+P_2\le C$, which is same to $E[SO_1]=E[SO_2]$, but this constraint was not included or reflected in the model of G(2010). Besides, G(2010) didn't discriminate the regions [8], [9] and [7] in fig.1. That is to say, G (2010) treat the condition happened in region [8] and [9] as the same in region [7] in the computation the expected number of stockouts, as shown in the equation (C.4) and (C.5) in the appendix C, which will cause error reasoning in solving the non-programming problem. The reason for this is as following:

When the demand happened in the region [8] or [9], all the finished and postponed inventory will all be exerted where $P_1 + P_2 = C$, and this is basis to compute $E[SO_i]$. At the same time, the equation will decide how the postponed capacity (or generic products) is allocated to each type of final products, but as shown in G&M(2002), when the demand happened in region [8] or [9], one of the product demand is much more larger than the other one, so even all the postponement capacity is allocated to complete the product of large demand, the fill rate can't be raised to equal that of lower product demand. For example, in region [8], demand for product 2 is much more lager than product 1, so all the postponed capacity is allocated to product 2 to attempt to equalize the fill rate, as a result, the boundary function between [7] and [8] is: $\frac{S_1}{x_1} = \frac{S_2 + C}{x_2}$, similarly, the boundary function between [7] and [9] is $\frac{S_1 + C}{x_1} = \frac{S_2}{x_2}$, so we can get the expected number of stockouts for product 1 in region [7]-[9]:

$$E[SO_1\{7\}] = \int_{x_2=S_2}^{S_2+C} \int_{x_1=S_1+S_2+C-x_2}^{(S_1+C)x_2/S_2} (x_1 - S_1 - P_1) f(x_1, x_2) dx_1 dx_2 \ , \ E[SO_1\{8\}] = \int_{x_2=S_2+C}^{\infty} \int_{x_1=S_1}^{S_1 x_2/(S_2+C)} (x_1 - S_1) f(x_1, x_2) dx_1 dx_2 \ ,$$

$$E[SO_1\{9\}] = \int_{x_2=S_2}^{\infty} \int_{x_1=(S_1+C)x_2/S_2}^{\infty} (x_1 - S_1 - C) f(x_1, x_2) dx_1 dx_2$$

The equation for product 2, $E[SO_2\{7\}]$, $E[SO_2\{8\}]$ and $E[SO_2\{9\}]$ can be denoted by the same reason. But not the general expression (C.4) and (C.5) in the appendix C in G(2010), which will ignore the condition in region [8] and [9] in the process of computation which will influence the result of program solving.

3. The new model including penalty cost parameter

After the penalty cost parameter is introduced to substitute for the fill rate, the fill rate constraints can be relaxed, the expression and computation of $E[SO_i]$ can be simplified. (1) When none of products stock out, the generic inventory will not be used, and some finished inventory and generic inventory will be left over in the end of period, as the demands in region [1] in fig.1(b). (2) When only one of finished products stock outs, i.e. the demands in region [2]-[5], some or all generic inventory will be allocated to the product of shortage. (3) When both demands can't be met from the finished inventory directly in region [6], but $x_1 + x_2 \le S_1 + S_2 + C$, so the postponed inventory can be allocated as: $x_1 - S_1$ for product 1 and $x_2 - S_2$ for product 2. Obviously, there is none shortage of product in region [1]-[6], so there is no penalty cost when demands happened in these regions. (4) When all the postponed inventory are exhausted, both of the demands can't be met at all, as the condition in region [7]-[9] in fig.1(a) or the region [7] in fig.1(b). The allocation principle is that the demand of larger penalty cost will be all satisfied in the first place, attempting to minimize the total cost, so there is no need to discriminate the expression of $E[SO_j]$ which is different from G&M(2002) and G(2010). The reason behind this is simple, as long as the product can reduce the cost to more extent, the postponed inventory will be allocated to it, even all the generic inventory will be allocated to the product of $\max\{t_1, t_2\}$.

The total cost includes assembly labor and material cost, postponement cost, packaging cost, holding cost of finished product and generic product, and shortage cost, so the objective function is

$$E[TC_{PP}] = m \times (S_1 + S_2) + m \times C + w \times C + d \times (S_1 + S_2) + d \times E[P_1 + P_2] + h_F \times E[LO_1 + LO_2] + h_P \times LO_C + t_1 \times E[SO_1] + t_2 \times E[SO_2]$$

$$= (m+d) \times (S_1 + S_2) + (m+w+h_P) \times C + (d - h_P) \times E[P_1 + P_2] + h_F \times E[LO_1 + LO_2] + t_1 \times E[SO_1] + t_2 \times E[SO_2]$$

$$C, S_i \ge 0, \forall i.$$

$$E[P_1 + P_2] = \sum_{i=1}^{2} \int_{x_i=0}^{S_i} [\int_{x_{3-i}=S_{3-i}}^{S_{3-i}+C} (x_{3-i} - S_{3-i}) f(x_1, x_2) + \int_{x_{3-i}=S_{3-i}+C}^{\infty} C f(x_1, x_2)] dx_1 dx_2$$
$$+ \int_{x_1=S_1}^{S_1+C} \int_{x_2=S_2}^{C+S_1+S_2-x_1} (x_1 + x_2 - S_1 - S_2) f(x_1, x_2)] dx_1 dx_2 + \sum_{i=1}^{2} \int_{x_i=S_i}^{\infty} [\int_{x_{3-i}=S_{3-i}}^{\infty} (C \times \max\{0, sig(t_i - t_{3-i})\}) f(x_1, x_2)] dx_1 dx_2$$

$$E[LO_1+LO_2]=\sum_{i=1}^{2}[\int_{x_i=0}^{S_i}\int_{x_{3-i}=0}^{S_{3-i}}(C)f(x_i,x_{3-i})dx_i dx_{3-i}+\int_{x_i=0}^{S_i}\int_{x_{3-i}=S_{3-i}}^{S_{3-i}}(C+S_{3-i}-x_{3-i})f(x_i,x_{3-i})dx_i dx_{3-i}+$$
$$\int_{x_i=S_i}^{S_i+C}\int_{x_{3-i}=S_{3-i}}^{S_i+S_{3-i}+C-x_i}(C+S_i+S_{3-i}-x_i-x_{3-i})f(x_i,x_{3-i})dx_i dx_{3-i}]$$
$$E[SO_i]=\sum_{i=1}^{2}\int_{x_{3-i}=0}^{S_{3-i}}\int_{x_i=S_i}^{\infty}(x_i-S_i-C)f(x_i,x_{3-i})dx_i dx_{3-i}+\int_{x_{3-i}=S_{3-i}}^{\infty}\int_{x_i=S_i}^{\infty}[x_i-S_i-C\times\max\{0,sig(t_i-t_{3-i})\}]f(x_i,x_{3-i})dx_i dx_{3-i}$$
$$-\int_{x_{3-i}=S_{3-i}}^{S_{3-i}+c}\int_{x_i=S_i}^{S_i+S_{3-i}+C-x_{3-i}}[x_i-S_i-C\times\max\{0,sig(t_i-t_{3-i})\}]f(x_i,x_{3-i})dx_i dx_{3-i}$$

There are three decision variables S_1, S_2, C and only some non-negativity constraints while there are five varibles and more additional constraints in G(2010). Besides, the model can be solved by genetic algorithm. So the computation process will be simplified and thereby it will be easy to obersve the effects of variable C on the value of S_i and the total cost, when the postponed capacity is set to different value.

4. Conclusion

In future research, we can still consider two final products which are still assembled or customized from a common inventory or product platform, but the two products can be partial substituted for each other, such that one of the product with better characteristics can be used to substitute for the other one to meet demand, but the reverse substitution can't be accepted or price of the product with better characteristics is general higher than the other one and the substitution is relative to tradeoff between product price and fill rate of customer demand.

Acknowledgements

This work was supported by Natural Science Foundation of Chongqing, China under Grant No. KJ100412.

References

[1] Graman GA, Magazine MJ. A numerical analysis of capacitated postponement. *Production and Operations Management*, 2002;**11(3)**: 340－357.

[2] Graman GA. A partial-postponement decision cost model. *European Journal of Operational Research*, 2010;**201**:34-44.

[3] Wong H, Wikner J, Naim M. Analysis of form postponement based on optimal positioning of the differentiation point and stocking decisions. *International Journal of Production Research*, 2009; **47**: 1201-1224.

[4] Wong H, Potter A, Naim M: Evaluation of postponement in a soluble coffee supply chain: A case study. *International Journal of Production Economics*, in press, 2010.

[5] Lee HL, Tang CS. Modeling the costs and benefits of delayed product differentiation. *Management Science*, 1997;**43**: 40－53.

[6] Li S, Ragu-Nathan B, Ragu-Nathan TS. The impact of supply chain management practices on competitive advantage and organizational performance. *Omega*, 2007;**34**: 107–124.